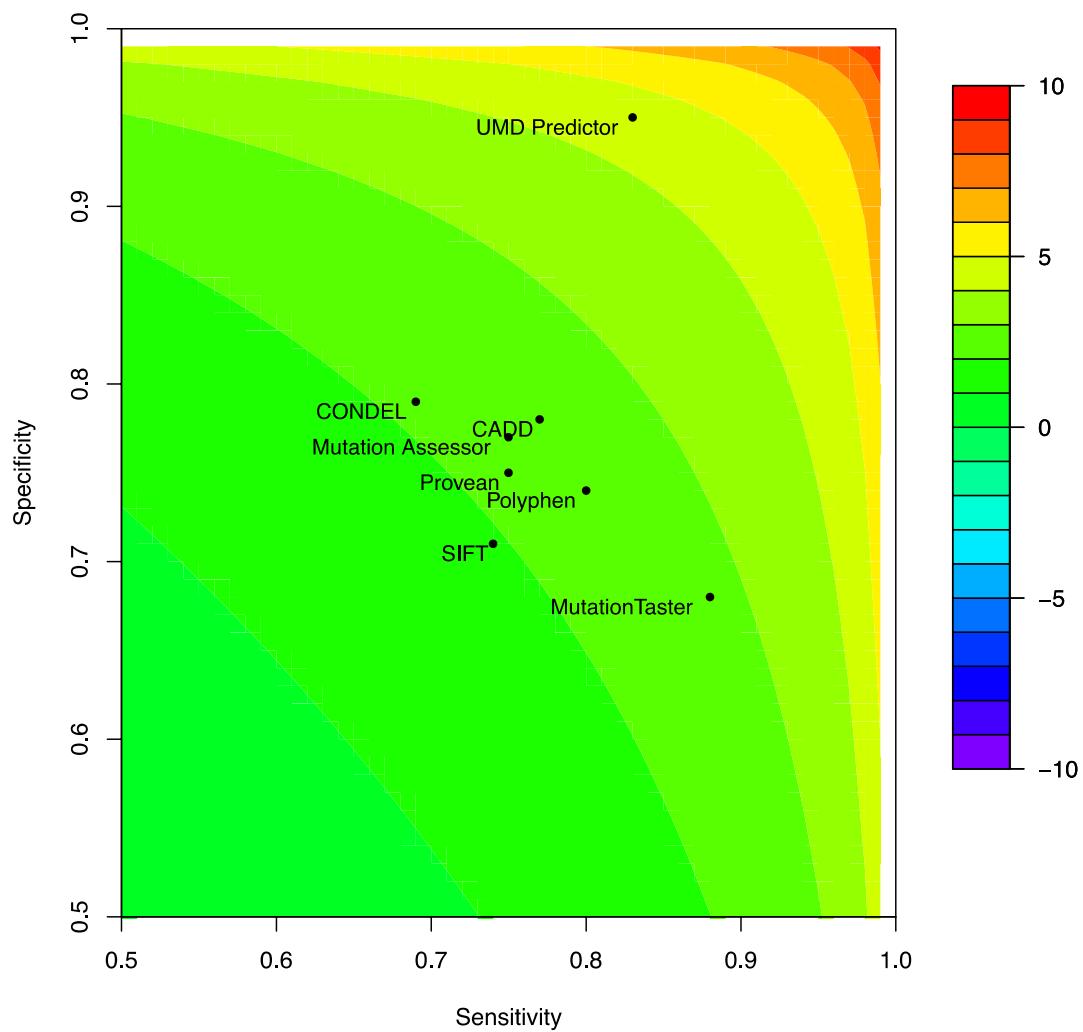
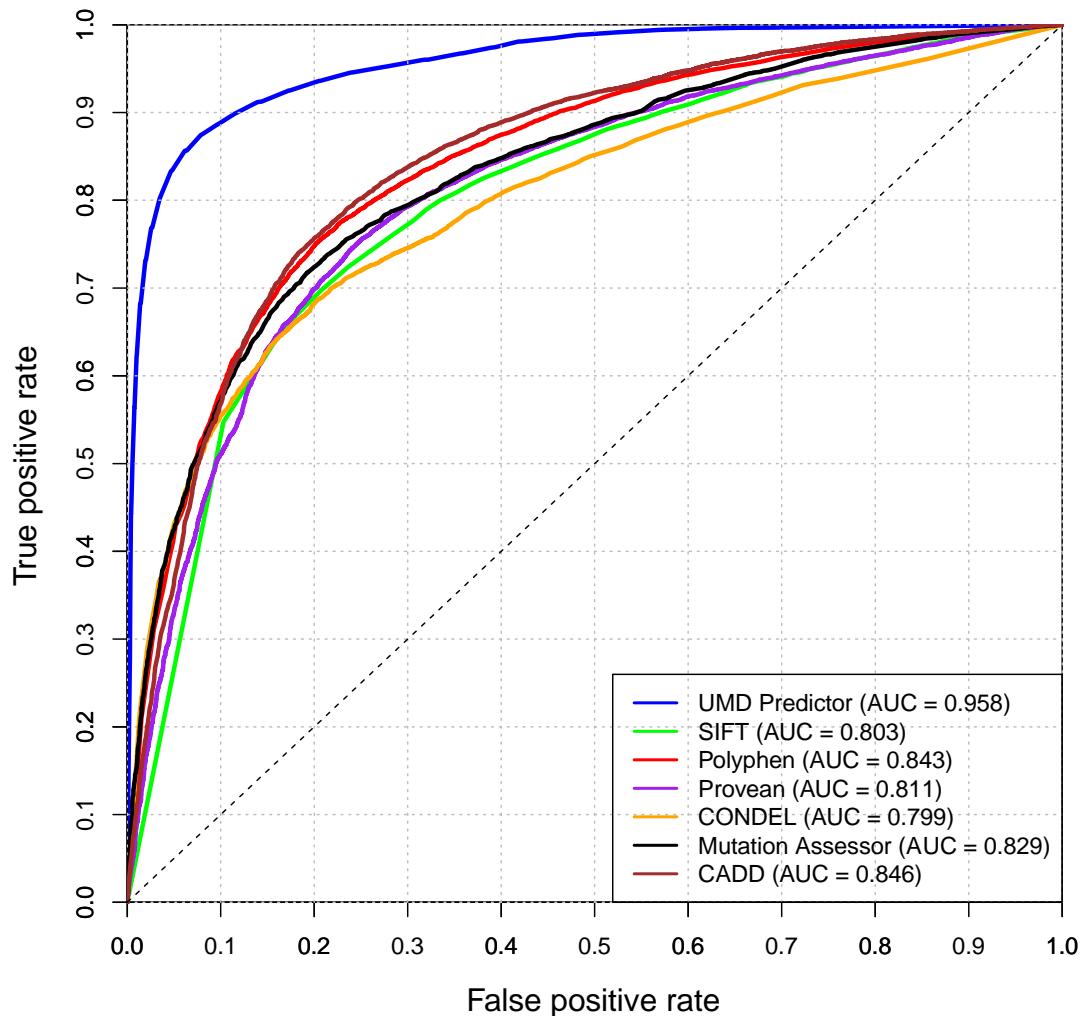


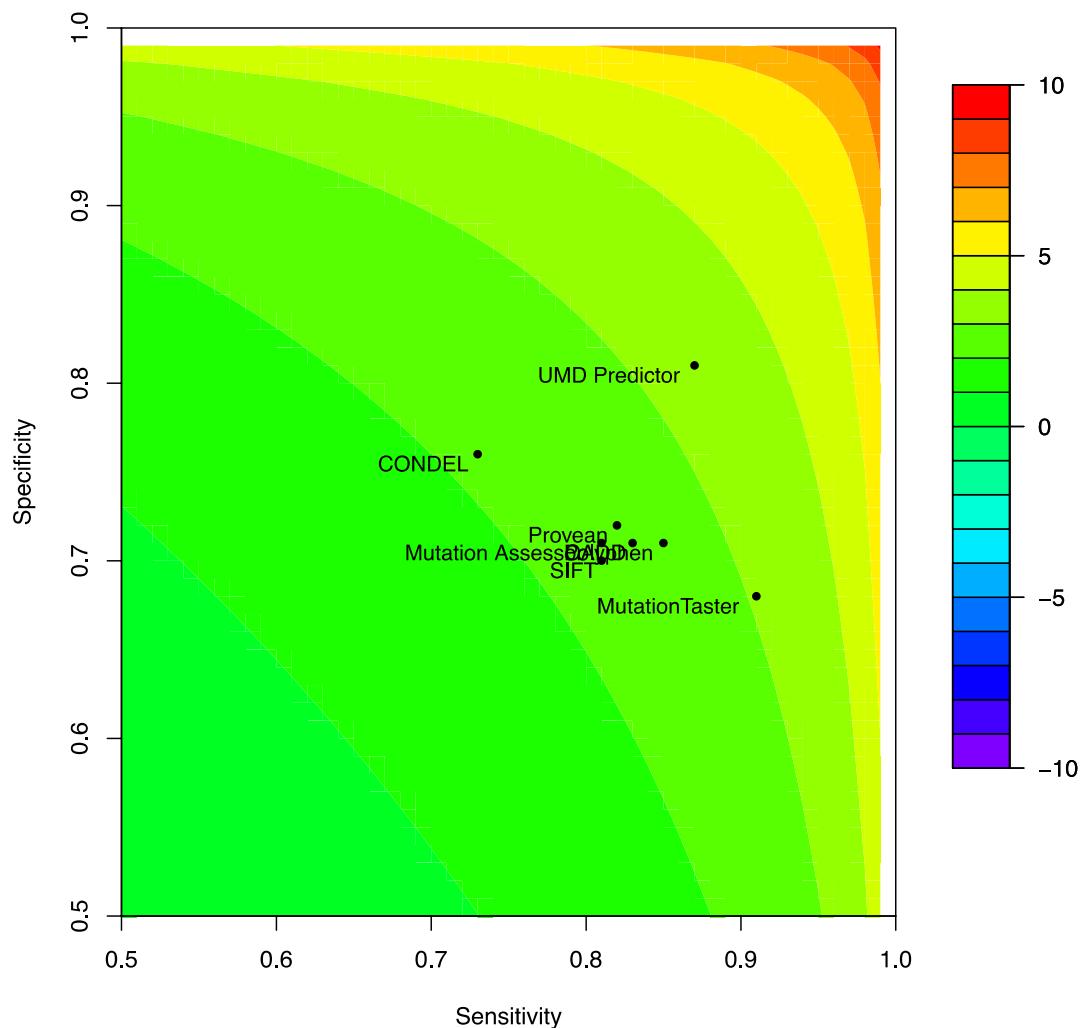
Supp. Figure S1: Multiple alignments of the Glutamic acid residue (E) at position 1003 of the *FBN1* gene transcript ENST00000316623. Red square: surrounding windows of 3 residues used to remove species with more than one substitution per window. Thus Mouse, Rat, and all species below Armadillo were excluded. The remaining 55 species were used to compute the conservation score of 97.68. This score was subsequently normalized based on the number and categories of remaining species to 71.9. In a second step, the conservation score is adjusted based on the Grantham substitution matrix in comparison to natural observed variants if any are available. In this example, one natural variant was observed for the Armadillo species: Glu>Ala with a Grantham score of 107. Therefore, all potential substitutions were thus compared to this natural variant. Thus a Glu>Lys mutation (Grantham score of 56) will result in a conservation score of 66.8, while a Glu>Val mutation (Grantham score of 121) will result in a conservation score of 73.3.



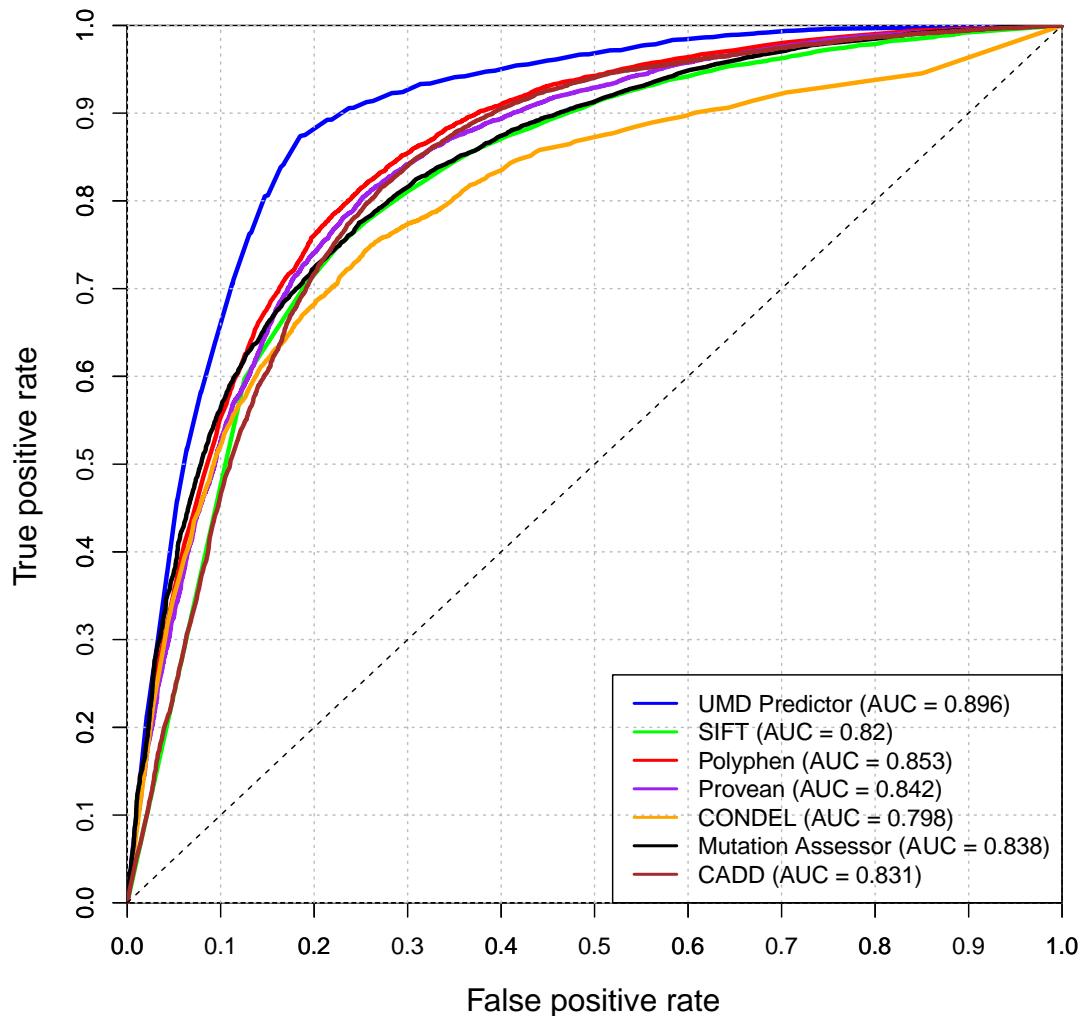
Supp. Figure S2: log(DOR) comparison between predictors using the full Varibench – dbSNP (Sherry et al. 2001; Sasidharan Nair and Vihinen 2013) full dataset (n=27.233). X-axis: sensitivity; Y-axis: specificity; color-coded scale: log(DOR).



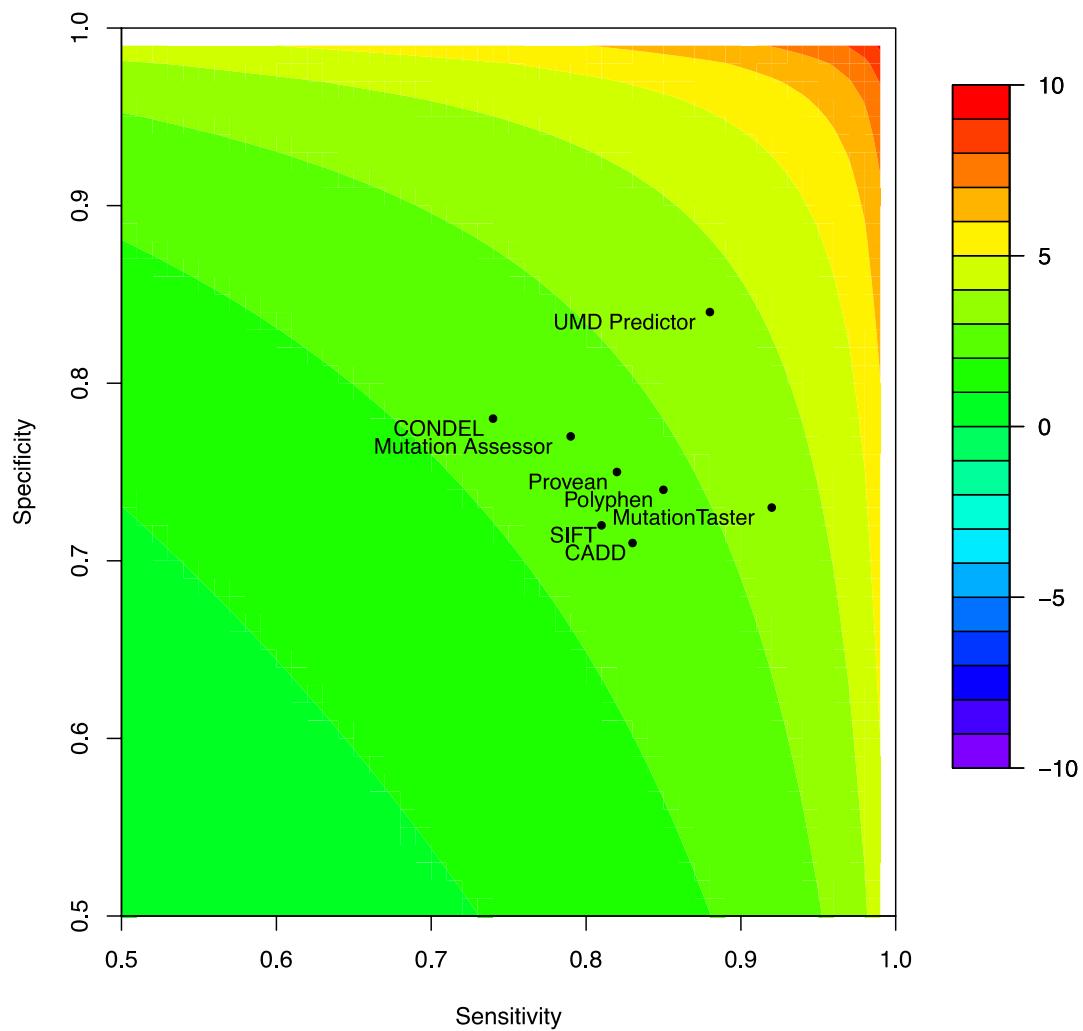
Supp. Figure S3: Sensitivity of methods in distinguishing pathogenic and non-pathogenic variants using the full Varibench - dbSNP (Sherry et al. 2001; Sasidharan Nair and Vihinen 2013) full dataset (n=27.233). Receiver Operating Characteristics (ROCs) are shown discriminating pathogenic mutations from non-pathogenic mutations defined by the Varibench - dbSNP (Sherry et al. 2001; Sasidharan Nair and Vihinen 2013) full dataset (n=27.233).



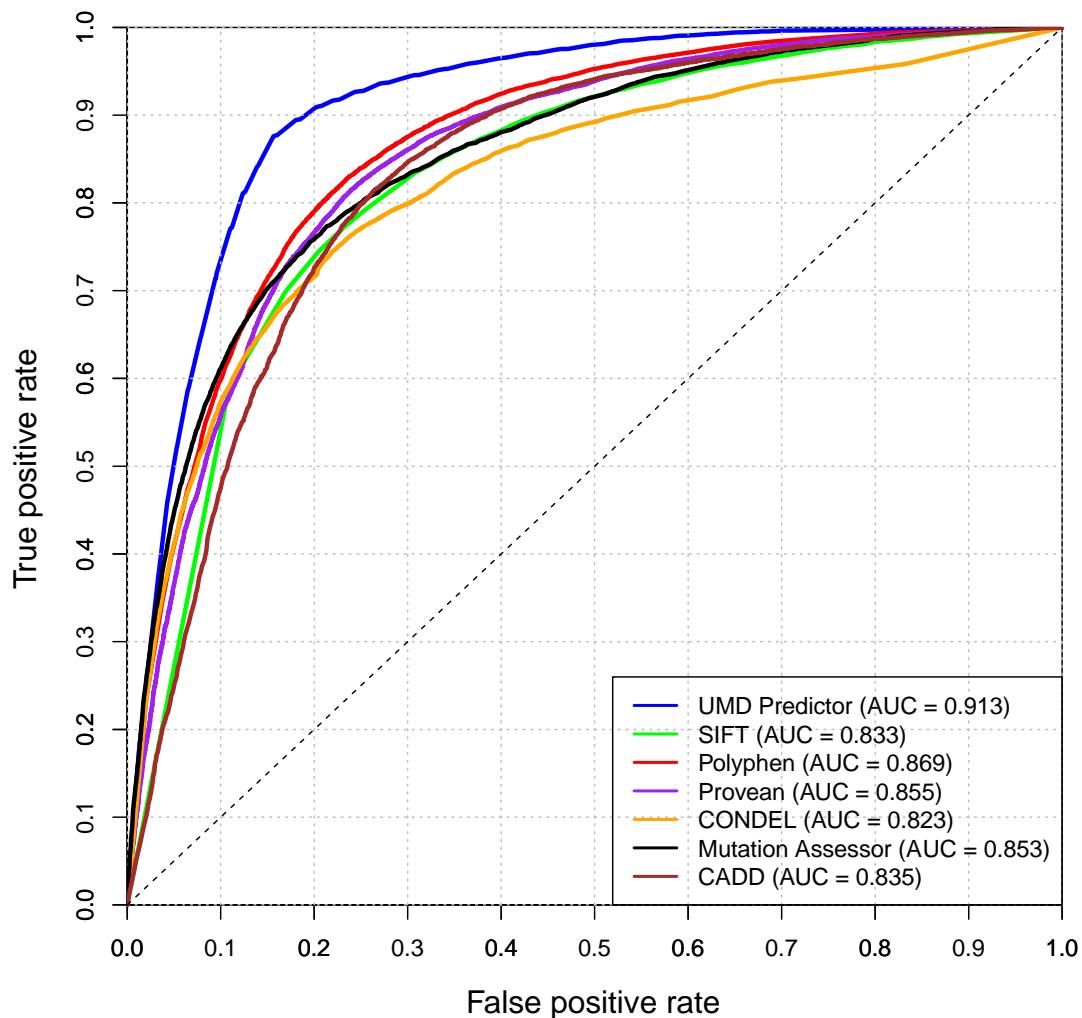
Supp. Figure S4: $\log(\text{DOR})$ comparison between predictors using the common Uniprot (UniProt Consortium 2014) dataset ($n=18.401$). X-axis: sensitivity; Y-axis: specificity; color-coded scale: $\log(\text{DOR})$.



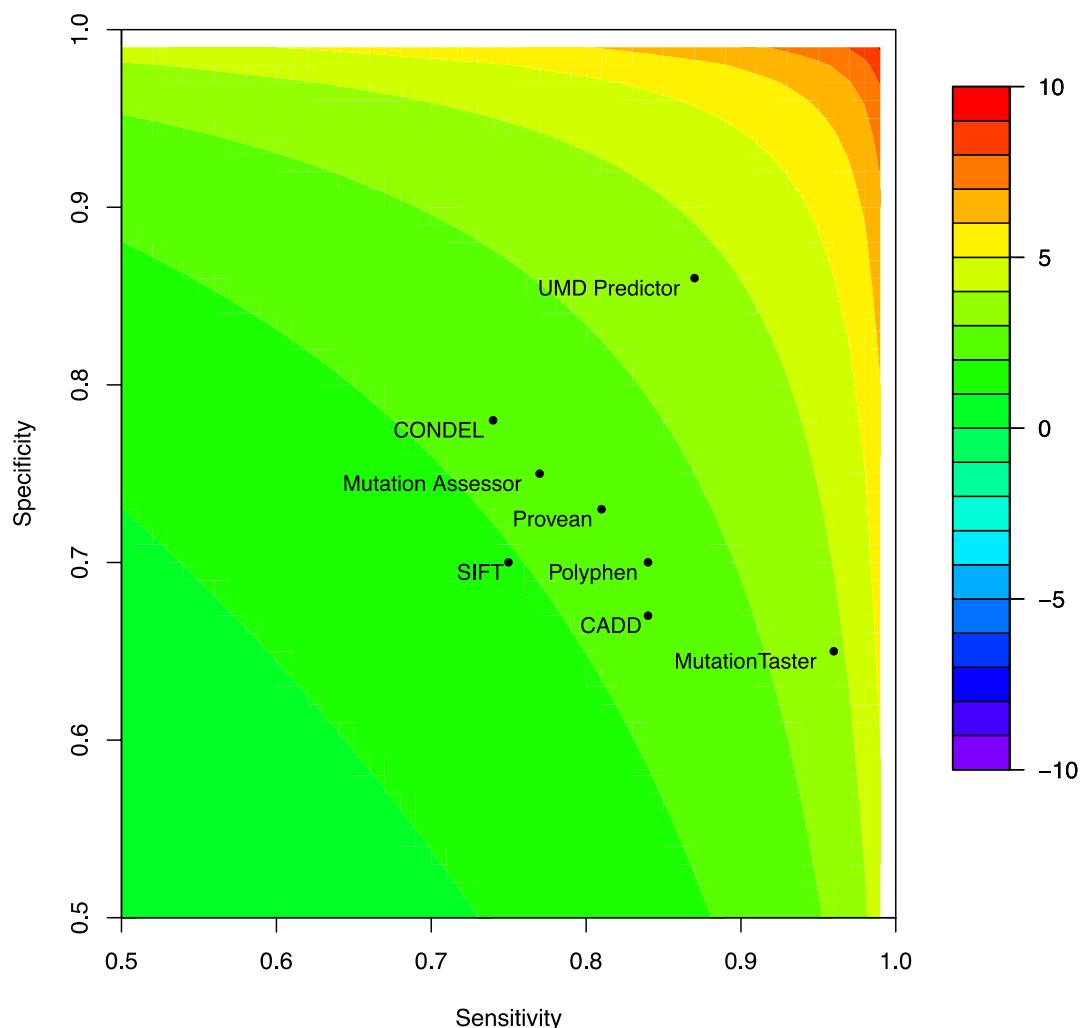
Supp. Figure S5: Sensitivity of methods in distinguishing pathogenic and non-pathogenic variants using the common Uniprot (UniProt Consortium 2014) dataset (n=18.401). Receiver Operating Characteristics (ROCs) are shown discriminating pathogenic mutations from non-pathogenic mutations defined by the common Uniprot (UniProt Consortium 2014) dataset (n=18.401).



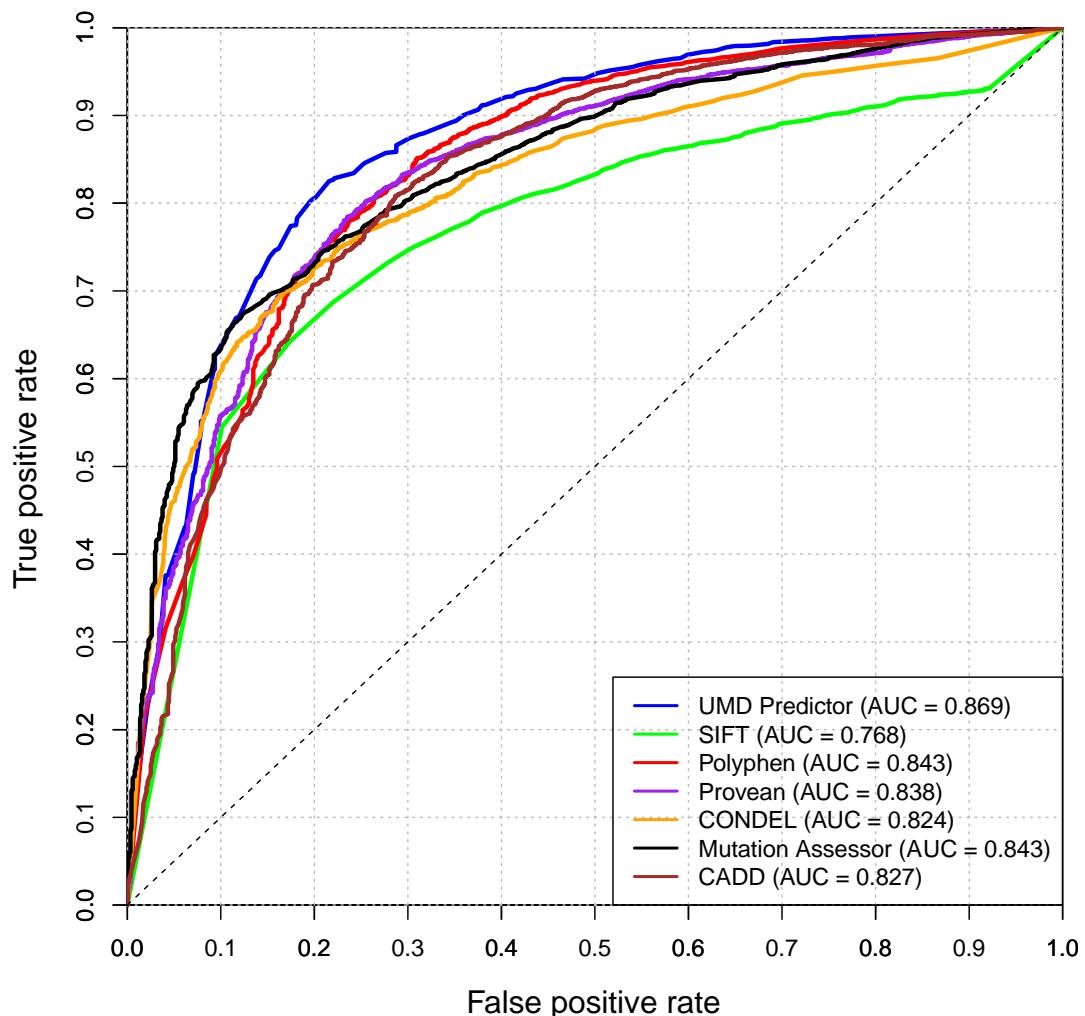
Supp. Figure S6: $\log(\text{DOR})$ comparison between predictors using the common Uniprot (UniProt Consortium 2014) full dataset ($n=57,646$). X-axis: sensitivity; Y-axis: specificity; color-coded scale: $\log(\text{DOR})$.



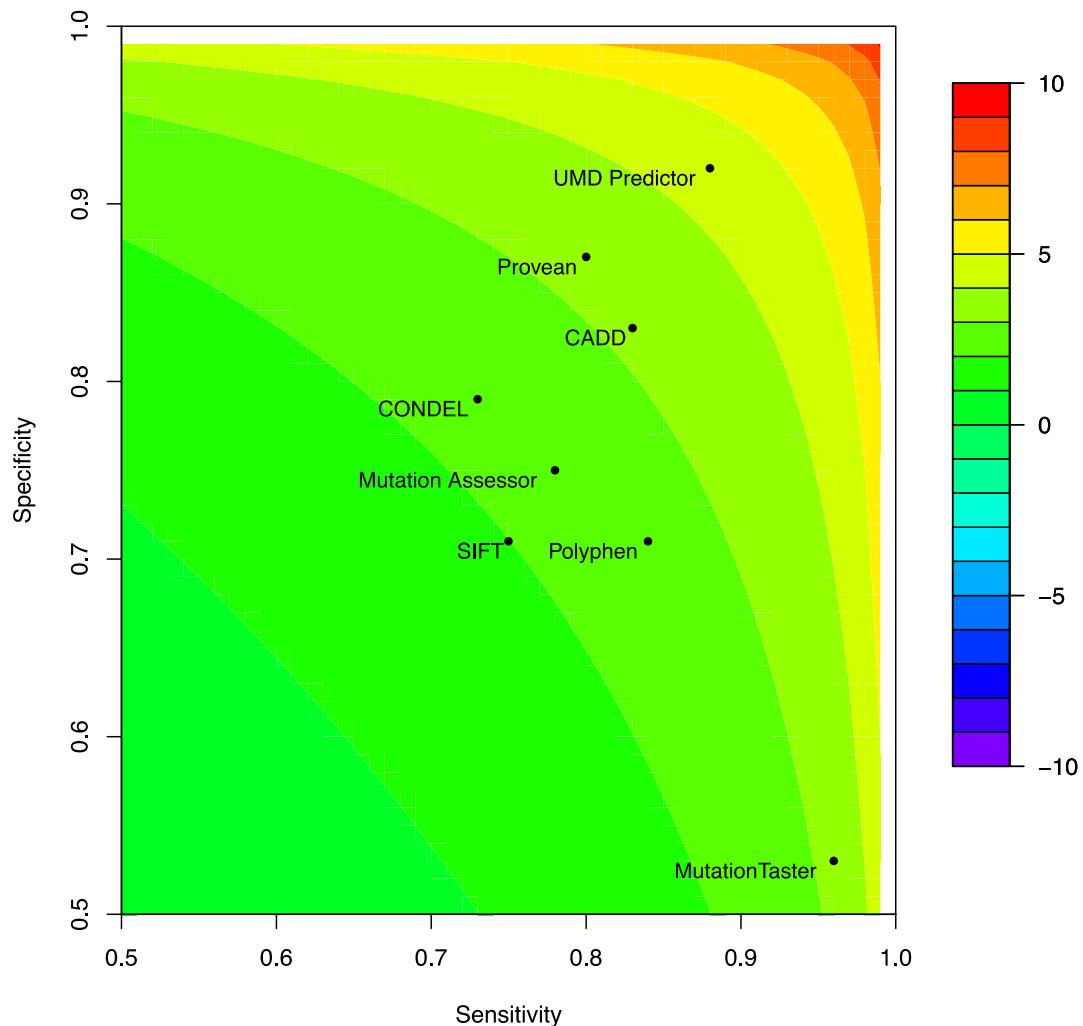
Supp. Figure S7: Sensitivity of methods in distinguishing pathogenic and non-pathogenic variants using the common Uniprot (UniProt Consortium 2014) full dataset (n=57,646). Receiver Operating Characteristics (ROCs) are shown discriminating pathogenic mutations from non-pathogenic mutations defined by the common Uniprot (UniProt Consortium 2014) full dataset (n=57,646).



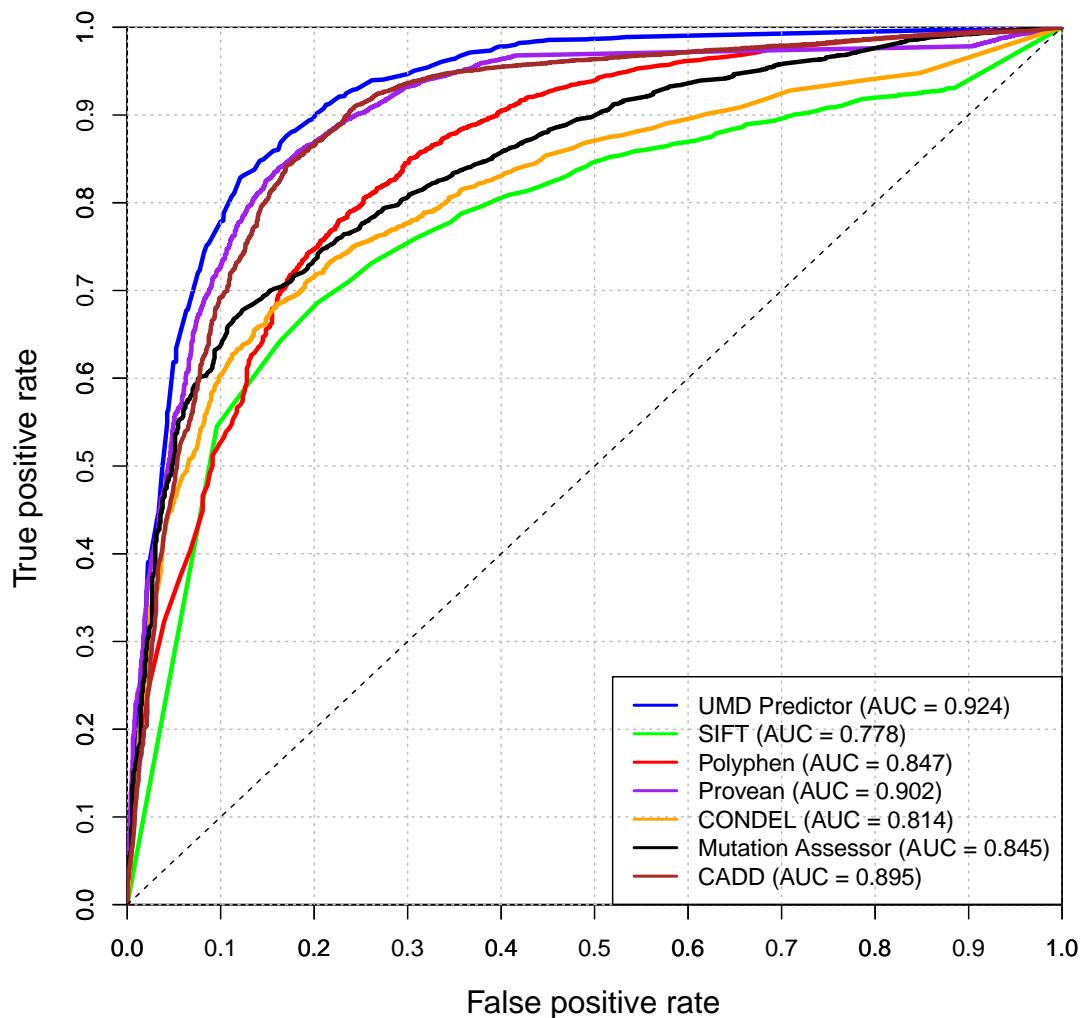
Supp. Figure S8: $\log(\text{DOR})$ comparison between predictors using the common Clinvar (Landrum et al. 2014) dataset ($n=10,023$). X-axis: sensitivity; Y-axis: specificity; color-coded scale: $\log(\text{DOR})$. Note that the MutationTaster2 tool does not perform any prediction on ClinVar pathogenic variations, but rather directly integrates conclusion from ClinVar.



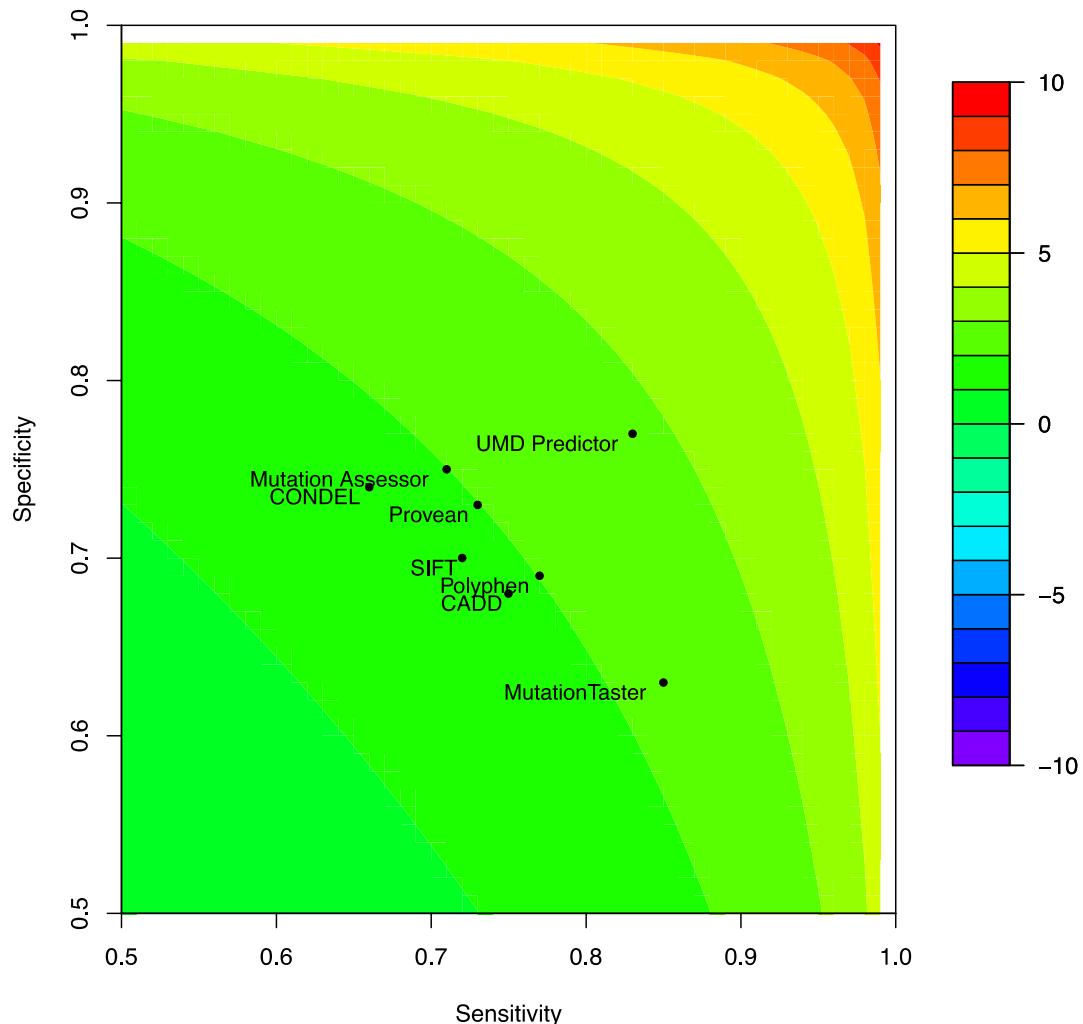
Supp. Figure S9: Sensitivity of methods in distinguishing pathogenic and non-pathogenic variants using the common Clinvar (Landrum et al. 2014) dataset (n=10.023). Receiver Operating Characteristics (ROCs) are shown discriminating pathogenic mutations from non-pathogenic mutations defined by the common Clinvar (Landrum et al. 2014) dataset (n=10.023).



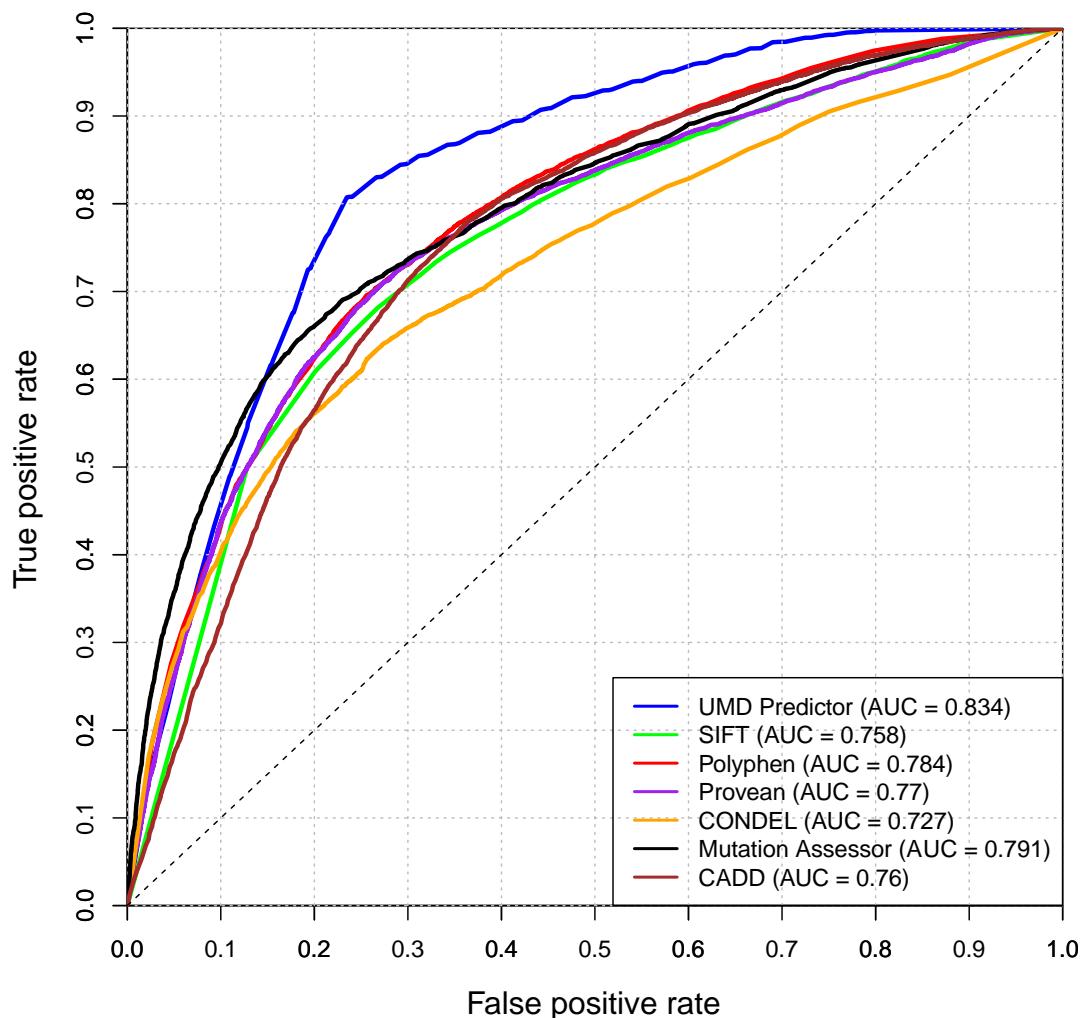
Supp. Figure S10: $\log(\text{DOR})$ comparison between predictors using the Clinvar (Landrum et al. 2014) full dataset ($n=12,486$). X-axis: sensitivity; Y-axis: specificity; color-coded scale: $\log(\text{DOR})$. Note that the MutationTaster2 tool does not perform any prediction on ClinVar pathogenic variations but rather directly integrates conclusion from ClinVar.



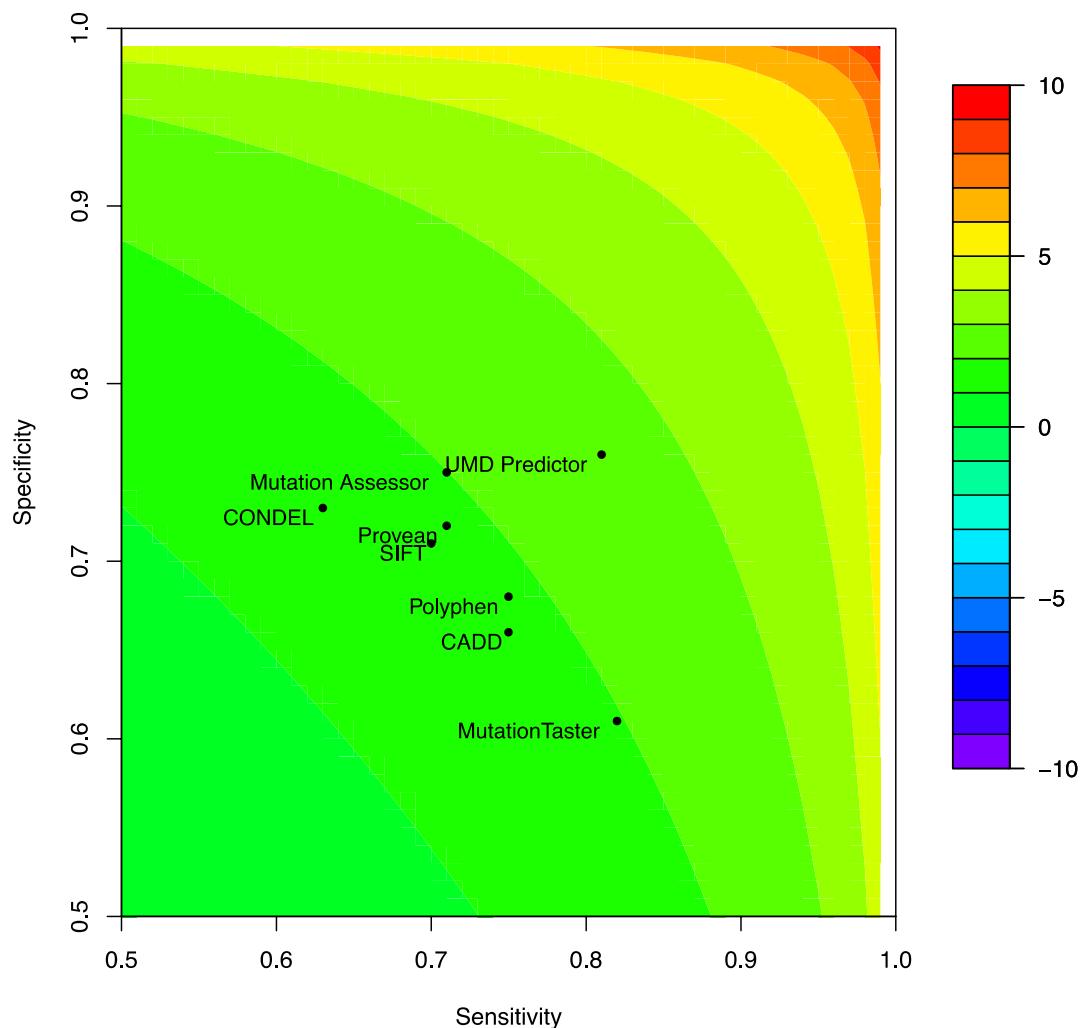
Supp. Figure S11: Sensitivity of methods in distinguishing pathogenic and non-pathogenic variants using the Clinvar (Landrum et al. 2014) full dataset (n=12.486).
Receiver Operating Characteristics (ROCs) are shown discriminating pathogenic mutations from non-pathogenic mutations defined by the Clinvar (Landrum et al. 2014) full dataset (n=12.486).



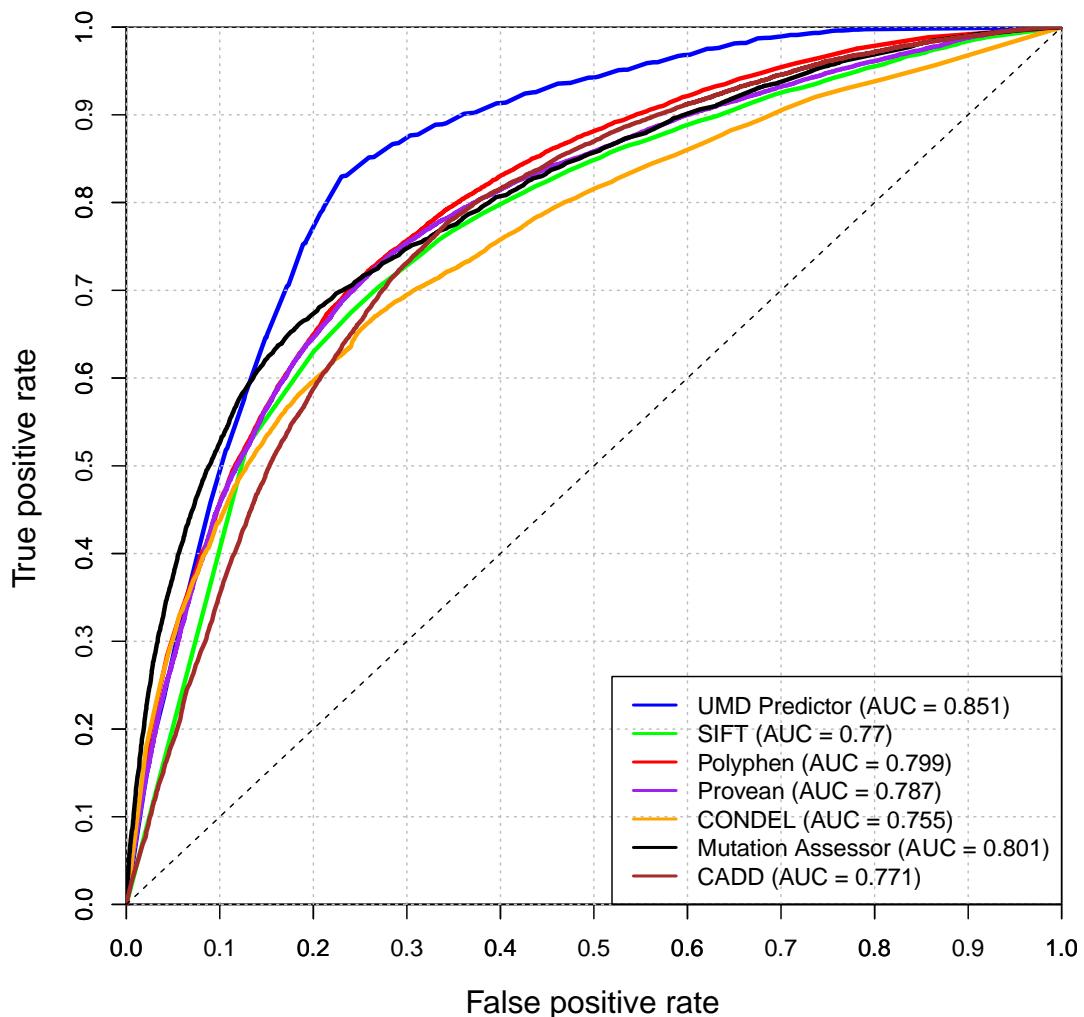
Supp. Figure S12: $\log(\text{DOR})$ comparison between predictors using the PredictSNP (Bendl et al. 2014) full dataset ($n=43.882$). X-axis: sensitivity; Y-axis: specificity; color-coded scale: $\log(\text{DOR})$. Note that the MutationTaster2 tool does not perform any prediction on ClinVar pathogenic variations but rather directly integrates conclusion from ClinVar.



Supp. Figure S13: Sensitivity of methods in distinguishing pathogenic and non-pathogenic variants using the PredictSNP (Bendl et al. 2014) full dataset (n=43.882).
Receiver Operating Characteristics (ROCs) are shown discriminating pathogenic mutations from non-pathogenic mutations defined by the PredictSNP (Bendl et al. 2014) full dataset (n=43.882).



Supp. Figure S14: $\log(\text{DOR})$ comparison between predictors using the common PredictSNP (Bendl et al. 2014) dataset ($n=24.300$). X-axis: sensitivity; Y-axis: specificity; color-coded scale: $\log(\text{DOR})$. Note that the MutationTaster2 tool does not perform any prediction on ClinVar pathogenic variations but rather directly integrates conclusion from ClinVar.



Supp. Figure S15: Sensitivity of methods in distinguishing pathogenic and non-pathogenic variants using the common PredictSNP (Bendl et al. 2014) dataset ($n=24.300$). Receiver Operating Characteristics (ROCs) are shown discriminating pathogenic mutations from non-pathogenic mutations defined by the common PredictSNP (Bendl et al. 2014) dataset ($n=24.300$).

Supp. Table S1: Comparison between UMD-Predictor and other prediction tools using the Varibench – dbSNP (Sherry et al. 2001; Sasidharan Nair and Vihinen 2013) full dataset (n=27.233)

	SIFT	PPH2	Provean	Mutation Assessor	CONDEL	Mutation Taster	CADD	UMD-Predictor
TP	13524	14806	14178	13525	13006	16423	10734	15570
TN	5403	5401	5758	5288	5955	4403	4642	7347
FP	2158	1932	1897	1582	1596	2028	1298	367
FN	4639	3708	4670	4485	5774	2295	3132	3107
PPV	0.86	0.88	0.88	0.90	0.89	0.89	0.89	0.98
NPV	0.54	0.59	0.55	0.54	0.51	0.66	0.60	0.70
Sensitivity	0.74	0.80	0.75	0.75	0.69	0.88	0.77	0.83
Specificity	0.71	0.74	0.75	0.77	0.79	0.68	0.78	0.95
Accuracy	0.74	0.78	0.75	0.76	0.72	0.83	0.78	0.87
MCC	0.43	0.51	0.47	0.48	0.44	0.55	0.52	0.73
DOR	7.0	11.4	9.0	10.0	8.4	15.6	11.9	92.8
Log(DOR)	1.94	2.43	2.20	2.31	2.13	2.75	2.47	4.53

TP. true positives; TN. true negatives; FP. false positives; FN. false negatives; PPV. positive predictive value; NPV. negative predictive value; MCC. Matthews correlation coefficient; DOR. diagnostic odds ratio.

Supp. Table S2: Receiver Operating Characteristics (ROCs) area for 7 predictors using the Varibench – dbSNP (Sherry et al. 2001; Sasidharan Nair and Vihinen 2013) full dataset (n=27.233)

	Confidence	ROC area	Standard error	Min ROC area	Max ROC area
UMD-Predictor	0.95	0.954	0.002	0.950	0.957
SIFT	0.95	0.784	0.005	0.774	0.794
PPH2_VAR	0.95	0.826	0.004	0.819	0.834
PROVEAN	0.95	0.789	0.004	0.781	0.797
CONDEL	0.95	0.778	0.004	0.771	0.786
MUT-ASS	0.95	0.813	0.004	0.806	0.820
CADD	0.95	0.834	0.004	0.827	0.841

Min ROC area = lower bound for the confidence interval of a vector of length two; Max ROC area = upper bound for the confidence interval of a vector of length two. All data were generated using the “ci.cvAUC” function of the “cvAUC” package (<https://github.com/ledell/cvAUC>) for the ROCR R-package (Sing et al. 2005).

Supp. Table S3: Comparison between UMD-Predictor and other prediction tools using the common Uniprot (UniProt Consortium 2014) dataset (n=18.401)

	SIFT	PPH2	Provean	Mutation Assessor	CONDEL	Mutation Taster	CADD	UMD-Predictor
TP	7947	8510	8243	8053	7274	9099	8330	8744
TN	5728	5940	6071	6004	6372	5712	5938	6827
FP	2462	2460	2331	2398	2029	2690	2464	1575
FN	1849	1489	1756	1946	2724	900	1669	1255
PPV	0.76	0.78	0.78	0.77	0.78	0.77	0.77	0.85
NPV	0.76	0.8	0.78	0.76	0.7	0.86	0.78	0.84
Sensitivity	0.81	0.85	0.82	0.81	0.73	0.91	0.83	0.87
Specificity	0.7	0.71	0.72	0.71	0.76	0.68	0.71	0.81
Accuracy	0.76	0.79	0.78	0.76	0.74	0.8	0.78	0.85
MCC	0.52	0.57	0.55	0.52	0.48	0.61	0.55	0.69
DOR	9.9	13.9	11.7	10.4	8.6	21.5	12.0	28.5
Log(DOR)	2.30	2.63	2.46	2.35	2.15	3.07	2.48	3.35

TP. true positives; TN. true negatives; FP. false positives; FN. false negatives; PPV. positive predictive value; NPV. negative predictive value; MCC. Matthews correlation coefficient; DOR. diagnostic odds ratio.

Supp. Table S4: Receiver Operating Characteristics (ROCs) area for 7 predictors using the common Uniprot (UniProt Consortium 2014) dataset (n=18.401)

	Confidence	ROC area	Standard error	Min ROC area	Max ROC area
UMD-Predictor	0.95	0.896	0.003	0.891	0.901
SIFT	0.95	0.820	0.004	0.813	0.828
PPH2_VAR	0.95	0.853	0.003	0.847	0.859
PROVEAN	0.95	0.842	0.003	0.836	0.848
CONDEL	0.95	0.798	0.003	0.791	0.805
MUT-ASS	0.95	0.838	0.003	0.832	0.844
CADD	0.95	0.831	0.003	0.825	0.837

Min ROC area = lower bound for the confidence interval of a vector of length two; Max ROC area = upper bound for the confidence interval of a vector of length two. All data were generated using the “ci.cvAUC” function of the “cvAUC” package (<https://github.com/ledell/cvAUC>) for the ROCR R-package (Sing et al. 2005).

Supp. Table S5: Comparison between UMD-Predictor and other prediction tools using the common Uniprot (UniProt Consortium 2014) full dataset (n=57.646)

	SIFT	PPH2	Provean	Mutation Assessor	CONDEL	Mutation Taster	CADD	UMD-Predictor
TP	15840	17668	15235	15546	14353	18441	8767	17664
TN	22943	27100	20300	24635	23614	24410	6550	28309
FP	8721	9609	6764	7400	6579	8927	2631	5343
FN	3746	3146	3268	4219	4979	1633	1754	2480
PPV	0.64	0.65	0.69	0.68	0.69	0.67	0.77	0.77
NPV	0.86	0.9	0.86	0.85	0.83	0.94	0.79	0.92
Sensitivity	0.81	0.85	0.82	0.79	0.74	0.92	0.83	0.88
Specificity	0.72	0.74	0.75	0.77	0.78	0.73	0.71	0.84
Accuracy	0.76	0.78	0.78	0.78	0.77	0.8	0.78	0.85
MCC	0.52	0.56	0.56	0.54	0.52	0.63	0.55	0.70
DOR	11.0	16.1	13.7	12.6	10.1	31.1	12.0	38.5
Log(DOR)	2.39	2.78	2.61	2.53	2.31	3.44	2.48	3.65

TP. true positives; TN. true negatives; FP. false positives; FN. false negatives; PPV. positive predictive value; NPV. negative predictive value; MCC. Matthews correlation coefficient; DOR. diagnostic odds ratio.

Supp. Table S6: Receiver Operating Characteristics (ROCs) area for 7 predictors using the Uniprot (UniProt Consortium 2014) full dataset (n=57.646)

	Confidence	ROC area	Standard error	Min ROC area	Max ROC area
UMD-Predictor	0.95	0.913	0.001	0.911	0.916
SIFT	0.95	0.833	0.002	0.829	0.837
PPH2_VAR	0.95	0.869	0.002	0.866	0.872
PROVEAN	0.95	0.855	0.002	0.852	0.858
CONDEL	0.95	0.823	0.002	0.819	0.827
MUT-ASS	0.95	0.853	0.002	0.850	0.856
CADD	0.95	0.835	0.003	0.829	0.840

Min ROC area = lower bound for the confidence interval of a vector of length two; Max ROC area = upper bound for the confidence interval of a vector of length two. All data were generated using the “ci.cvAUC” function of the “cvAUC” package (<https://github.com/ledell/cvAUC>) for the ROCR R-package (Sing et al. 2005).

Supp. Table S7: Comparison between UMD-Predictor and other prediction tools using the common Clinvar (Landrum et al. 2014) dataset (n=10.023)

	SIFT	PPH2	Provean	Mutation Assessor	CONDEL	Mutation Taster	CADD	UMD-Predictor
TP	6475	7659	7392	7085	6783	8772	7717	8002
TN	508	609	643	654	624	571	583	750
FP	222	267	233	222	178	305	293	126
FN	2164	1488	1753	2062	2342	375	1430	1145
PPV	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.98
NPV	0.19	0.29	0.27	0.24	0.21	0.60	0.29	0.40
Sensitivity	0.75	0.84	0.81	0.77	0.74	0.96	0.84	0.87
Specificity	0.70	0.70	0.73	0.75	0.78	0.65	0.67	0.86
Accuracy	0.75	0.82	0.8	0.77	0.75	0.93	0.83	0.87
MCC	0.26	0.37	0.36	0.33	0.31	0.59	0.36	0.53
DOR	7.0	12.3	11.5	10.0	10.1	44.6	10.7	41.1
Log(DOR)	1.95	2.51	2.44	2.31	2.31	3.80	2.37	3.72

TP. true positives; TN. true negatives; FP. false positives; FN. false negatives; PPV. positive predictive value; NPV. negative predictive value; MCC. Matthews correlation coefficient; DOR. diagnostic odds ratio. Note that the MutationTaster2 tool does not perform any prediction on ClinVar pathogenic variations but rather directly integrates conclusion from ClinVar.

Supp. Table S8: Receiver Operating Characteristics (ROCs) area for 7 predictors using the common Clinvar (Landrum et al. 2014) dataset (n=10.023)

	Confidence	ROC area	Standard error	Min ROC area	Max ROC area
UMD-Predictor	0.95	0.869	0.007	0.855	0.883
SIFT	0.95	0.768	0.010	0.748	0.789
PPH2_VAR	0.95	0.843	0.008	0.828	0.858
PROVEAN	0.95	0.838	0.007	0.824	0.851
CONDEL	0.95	0.824	0.007	0.811	0.838
MUT-ASS	0.95	0.843	0.006	0.831	0.856
CADD	0.95	0.827	0.008	0.811	0.842

Min ROC area = lower bound for the confidence interval of a vector of length two; Max ROC area = upper bound for the confidence interval of a vector of length two. All data were generated using the “ci.cvAUC” function of the “cvAUC” package (<https://github.com/ledell/cvAUC>) for the ROCR R-package (Sing et al. 2005).

Supp. Table S9: Comparison between UMD-Predictor and other prediction tools using the Clinvar (Landrum et al. 2014) full dataset (n=12.486)

	SIFT	PPH2	Provean	Mutation Assessor	CONDEL	Mutation Taster	CADD	UMD-Predictor
TP	6641	8801	8445	7994	7619	9041	8861	9235
TN	615	656	1564	673	674	932	1510	1648
FP	249	274	242	226	183	837	301	152
FN	2239	1715	2098	2298	2859	396	1793	1317
PPV	0.96	0.97	0.97	0.97	0.98	0.92	0.97	0.98
NPV	0.22	0.28	0.43	0.23	0.19	0.70	0.46	0.56
Sensitivity	0.75	0.84	0.8	0.78	0.73	0.96	0.83	0.88
Specificity	0.71	0.71	0.87	0.75	0.79	0.53	0.83	0.92
Accuracy	0.74	0.83	0.81	0.77	0.73	0.89	0.83	0.88
MCC	0.29	0.37	0.52	0.32	0.29	0.55	0.53	0.65
DOR	7.3	12.9	26.8	10.6	10.2	27.1	23.8	84.3
Log(DOR)	1.99	2.55	3.29	2.36	2.32	3.30	3.17	4.43

TP. true positives; TN. true negatives; FP. false positives; FN. false negatives; PPV. positive predictive value; NPV. negative predictive value; MCC. Matthews correlation coefficient; DOR. diagnostic odds ratio. Note that the MutationTaster2 tool does not perform any prediction on ClinVar pathogenic variations but rather directly integrates conclusion from ClinVar.

Supp. Table S10: Receiver Operating Characteristics (ROCs) area for 7 predictors using the Clinvar (Landrum et al. 2014) full dataset (n=12,486)

	Confidence	ROC area	Standard error	Min ROC area	Max ROC area
UMD-Predictor	0.95	0.924	0.004	0.916	0.931
SIFT	0.95	0.778	0.010	0.759	0.796
PPH2_VAR	0.95	0.847	0.007	0.833	0.862
PROVEAN	0.95	0.902	0.004	0.894	0.910
CONDEL	0.95	0.814	0.007	0.801	0.827
MUT-ASS	0.95	0.845	0.006	0.833	0.857
CADD	0.95	0.895	0.004	0.886	0.904

Min ROC area = lower bound for the confidence interval of a vector of length two; Max ROC area = upper bound for the confidence interval of a vector of length two. All data were generated using the “ci.cvAUC” function of the “cvAUC” package (<https://github.com/ledell/cvAUC>) for the ROCR R-package (Sing et al. 2005).

Supp. Table S11: Comparison between UMD-Predictor and other prediction tools using the PredictSNP (Bendl et al. 2014) full dataset (n=43.882)

	SIFT	PPH2	Provean	Mutation Assessor	CONDEL	Mutation Taster	CADD	UMD-Predictor
TP	13843	14506	14352	13156	12050	16657	10524	16339
TN	14635	15248	17210	15335	15043	14218	11263	17498
FP	6126	6939	6378	5073	5150	8419	5245	5309
FN	5267	4419	5340	5312	6221	2928	3530	3340
PPV	0.69	0.68	0.69	0.72	0.70	0.66	0.67	0.75
NPV	0.74	0.78	0.76	0.74	0.71	0.83	0.76	0.84
Sensitivity	0.72	0.77	0.73	0.71	0.66	0.85	0.75	0.83
Specificity	0.7	0.69	0.73	0.75	0.74	0.63	0.68	0.77
Accuracy	0.71	0.72	0.73	0.73	0.70	0.73	0.71	0.80
MCC	0.43	0.45	0.46	0.46	0.41	0.49	0.43	0.60
DOR	6.0	7.5	7.3	7.3	5.5	9.6	6.4	16.3
Log(DOR)	1.79	2.01	1.99	1.99	1.71	2.27	1.85	2.79

TP. true positives; TN. true negatives; FP. false positives; FN. false negatives; PPV. positive predictive value; NPV. negative predictive value; MCC. Matthews correlation coefficient; DOR. diagnostic odds ratio. Note that the MutationTaster2 tool does not perform any prediction on ClinVar pathogenic variations but rather directly integrates conclusion from ClinVar.

Supp. Table S12: Receiver Operating Characteristics (ROCs) area for 7 predictors using the PredictSNP (Bendl et al. 2014) full dataset (n=43.882)

	Confidence	ROC area	Standard error	Min ROC area	Max ROC area
UMD-Predictor	0.95	0.834	0.003	0.829	0.840
SIFT	0.95	0.758	0.003	0.751	0.764
PPH2_VAR	0.95	0.784	0.003	0.778	0.790
PROVEAN	0.95	0.770	0.003	0.764	0.776
CONDEL	0.95	0.727	0.003	0.721	0.734
MUT-ASS	0.95	0.791	0.003	0.785	0.797
CADD	0.95	0.760	0.003	0.754	0.766

Min ROC area = lower bound for the confidence interval of a vector of length two; Max ROC area = upper bound for the confidence interval of a vector of length two. All data were generated using the “ci.cvAUC” function of the “cvAUC” package (<https://github.com/ledell/cvAUC>) for the ROCR R-package (Sing et al. 2005).

Supp. Table S13: Comparison between UMD-Predictor and other prediction tools using the common PredictSNP (Bendl et al. 2014) dataset (n=24.300)

	SIFT	PPH2	Provean	Mutation Assessor	CONDEL	Mutation Taster	CADD	UMD-Predictor
TP	8503	9085	8636	8558	7655	9944	9090	9788
TN	8622	8251	8789	9099	8950	7468	8082	9289
FP	3561	3931	3394	3084	3233	4715	4101	2894
FN	3614	3032	3481	3559	4462	2173	3027	2329
PPV	0.70	0.70	0.72	0.74	0.70	0.68	0.69	0.77
NPV	0.70	0.73	0.72	0.72	0.67	0.77	0.73	0.80
Sensitivity	0.70	0.75	0.71	0.71	0.63	0.82	0.75	0.81
Specificity	0.71	0.68	0.72	0.75	0.73	0.61	0.66	0.76
Accuracy	0.70	0.71	0.72	0.73	0.68	0.72	0.71	0.79
MCC	0.41	0.43	0.43	0.45	0.37	0.44	0.42	0.57
DOR	5.7	6.4	6.3	7.3	4.6	7.1	5.8	13.5
Log(DOR)	1.74	1.85	1.84	1.99	1.53	1.96	1.76	2.60

TP. true positives; TN. true negatives; FP. false positives; FN. false negatives; PPV. positive predictive value; NPV. negative predictive value; MCC. Matthews correlation coefficient; DOR. diagnostic odds ratio. Note that the MutationTaster2 tool does not perform any prediction on ClinVar pathogenic variations but rather directly integrates conclusion from ClinVar.

Supp. Table S14: Receiver Operating Characteristics (ROCs) area for 7 predictors using the common PredictSNP (Bendl et al. 2014) dataset (n=24.300)

	Confidence	ROC area	Standard error	Min ROC area	Max ROC area
UMD-Predictor	0.95	0.851	0.002	0.847	0.854
SIFT	0.95	0.770	0.003	0.765	0.775
PPH2_VAR	0.95	0.799	0.002	0.795	0.804
PROVEAN	0.95	0.787	0.002	0.783	0.791
CONDEL	0.95	0.755	0.003	0.750	0.760
MUT-ASS	0.95	0.801	0.002	0.797	0.806
CADD	0.95	0.771	0.003	0.766	0.777

Min ROC area = lower bound for the confidence interval of a vector of length two; Max ROC area = upper bound for the confidence interval of a vector of length two. All data were generated using the “ci.cvAUC” function of the “cvAUC” package (<https://github.com/ledell/cvAUC>) for the ROCR R-package (Sing et al. 2005).